SUPPLEMENTARY MATERIAL

Derivation of Alternative Model

In this section, we explore the parameter estimation for an alternative model. Specifically, letting $\mathcal{M}_i$ be the set of missing voxels of patch y_i, we treat $y_i^{\mathcal{M}_i}$ as latent variables, instead of explicitly modeling a low-dimensional representation x. We show the maximum likelihood updates of the model parameters under the likelihood (5). We employ the Expectation Conditional Maximization (ECM) [14], [18] variant of the Generalized Expectation Maximization, where parameter updates depend on the previous parameter estimates.

The complete data likelihood is

$$p(\mathcal{Y};\theta) = \prod_i \sum_k \pi_k \mathcal{N}(y_i^{\mathcal{O}_i}, y_i^{\mathcal{M}_i}; \mu_k^{\mathcal{O}_i}, \Sigma_k^{\mathcal{O}_i\mathcal{O}_i}). \tag{25}$$

The **expectation step** updates the statistics of the missing data, computed based on covariates of the known and unknown voxels:

$$\begin{aligned} \gamma_{ik} &\equiv \mathbf{E}[k_i] \\ &= \frac{\pi_k \mathcal{N}(y_i^{\mathcal{O}_i}; \mu_k^{\mathcal{O}_i}, \Sigma_k^{\mathcal{O}_i})}{\sum_k \pi_k \mathcal{N}(y_i^{\mathcal{O}_i}; \mu_k^{\mathcal{O}_i}, \Sigma_k^{\mathcal{O}_i})} \end{aligned} \tag{26}$$

$$\begin{aligned} \widehat{y}_{ij} &\equiv \mathbf{E}[y_{ij}] \\ &= \begin{cases} y_{ij} & \text{if } y_{ij} \text{ is observed} \\ \mu_{ij} + \Sigma_i^{j\mathcal{O}_i}(\Sigma_i^{\mathcal{O}_i\mathcal{O}_i})^{-1}(y_i^{\mathcal{O}_i} - \mu^{\mathcal{O}_i}) & \text{otherwise} \end{cases} \end{aligned} \tag{27}$$

$$\begin{aligned} \widehat{s}_{ijl} &\equiv \mathbf{E}\left[y_{ij}y_{il}\right] - \mathbf{E}[y_{ij}]\mathbf{E}[y_{il}] \\ &= \begin{cases} 0 & \text{if } y_{ij} \text{ or } y_{il} \text{ is observed} \\ \Sigma_i^{jl} - (\Sigma_i^{\mathcal{O}_i j})^T(\Sigma_i^{\mathcal{O}_i\mathcal{O}_i})^{-1}\Sigma_i^{\mathcal{O}_i l} & \text{otherwise} \end{cases} \end{aligned} \tag{28}$$

where the correction in $\widehat{s}_{ijl}$ can be interpreted as the uncertainty in the covariance estimation due to the missing values.

Given estimates for the missing data, the **maximization step** leads to familiar Gaussian Mixture Model parameters updates:

$$\mu_k = \frac{1}{\gamma_{ik}} \sum_i \gamma_{ik} \widehat{y}_{ik} \tag{29}$$

$$\Sigma_k = \frac{1}{\gamma_{ik}} \sum_i \gamma_{ik} \left[(\widehat{y}_{ik} - \mu_k)(\widehat{y}_{ik} - \mu_k)^T + S_i^T \right]. \tag{30}$$

$$\pi_k = \frac{1}{N} \sum_i \gamma_{ik} \tag{31}$$

where $[S_i]_{jl} = \widehat{s}_{ijl}$.

In additional to the latent missing voxels, we can still model each patch as coming from a low dimensional representation. We form $C_k = W_k W_k^T + \sigma_k^2 I$ as in (3), leading to the complete data likelihood:

$$p(\mathcal{Y};\theta) = \prod_i \sum_k \pi_k \mathcal{N}(y_i^{\mathcal{O}_i}, y_i^{\mathcal{M}_i}; \mu_k^{\mathcal{O}_i}, C_k^{\mathcal{O}_i\mathcal{O}_i}). \tag{32}$$

The **expectation steps** are then unchanged from (26)-(28) with C_k replacing Σ_k. The **maximization steps** are unchanged from (29)-(31), with Σ_k now the *empirical* covariance in (30). We let $U\Lambda V^T = \text{SVD}(\Sigma_k)$ be the singular value decomposition of Σ_k, leading to the low dimensional updates

$$\sigma_k^2 \leftarrow \frac{1}{d-q} \sum_{j=d+1}^{d} \Lambda(j,j) \tag{33}$$

$$W_k \leftarrow U(\Lambda - \sigma^2 I)^{1/2}. \tag{34}$$

Finally, we let $C_k = W_k W_k^T + \sigma_k^2 I$.

Unfortunately, both learning procedures involve estimating all of the missing voxel covariances, leading to a large and unstable optimization.

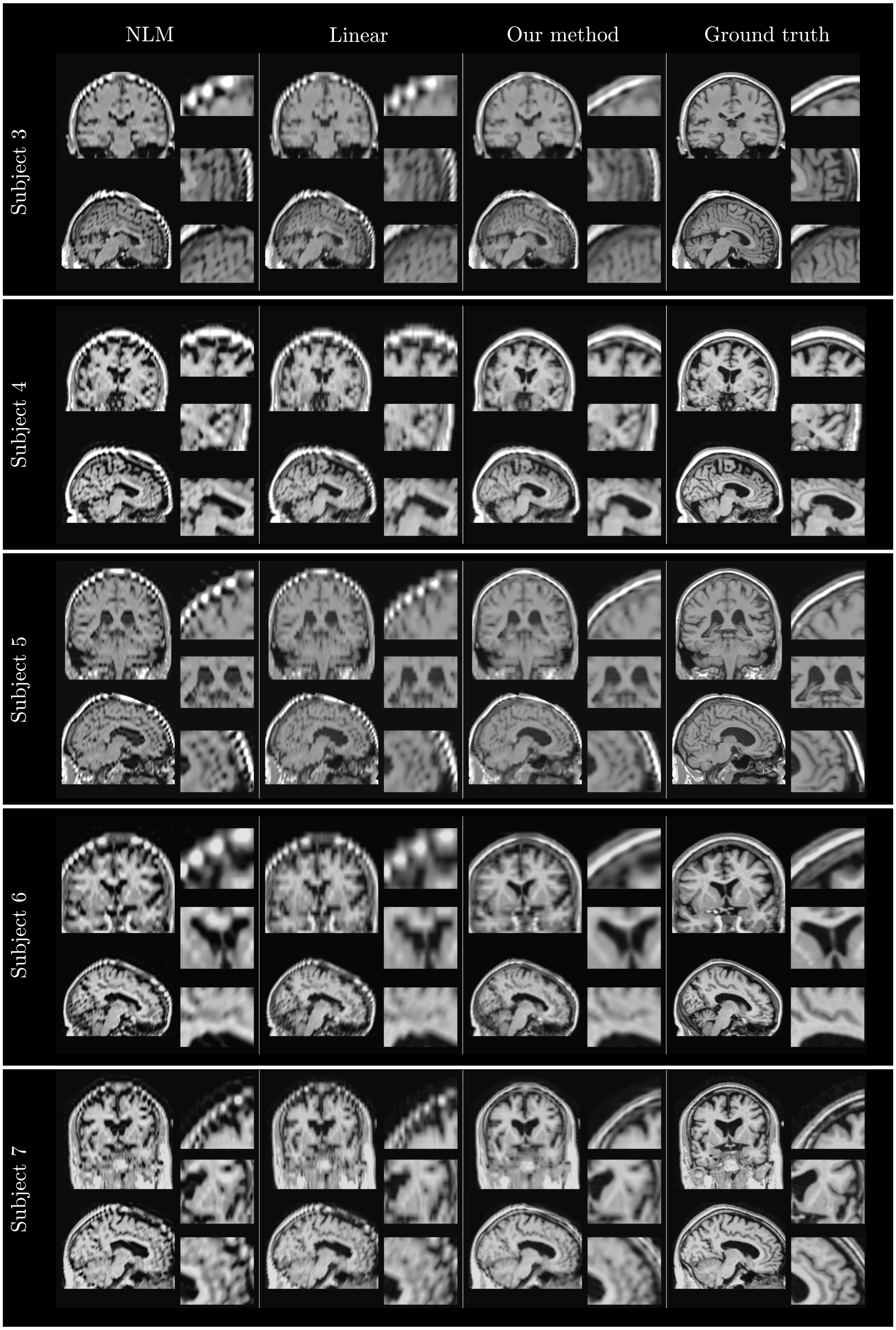

Fig. 11: Additional restorations in the ADNI dataset. Reconstruction by NLM, linear interpolation, and our method, and the original high resolution images.

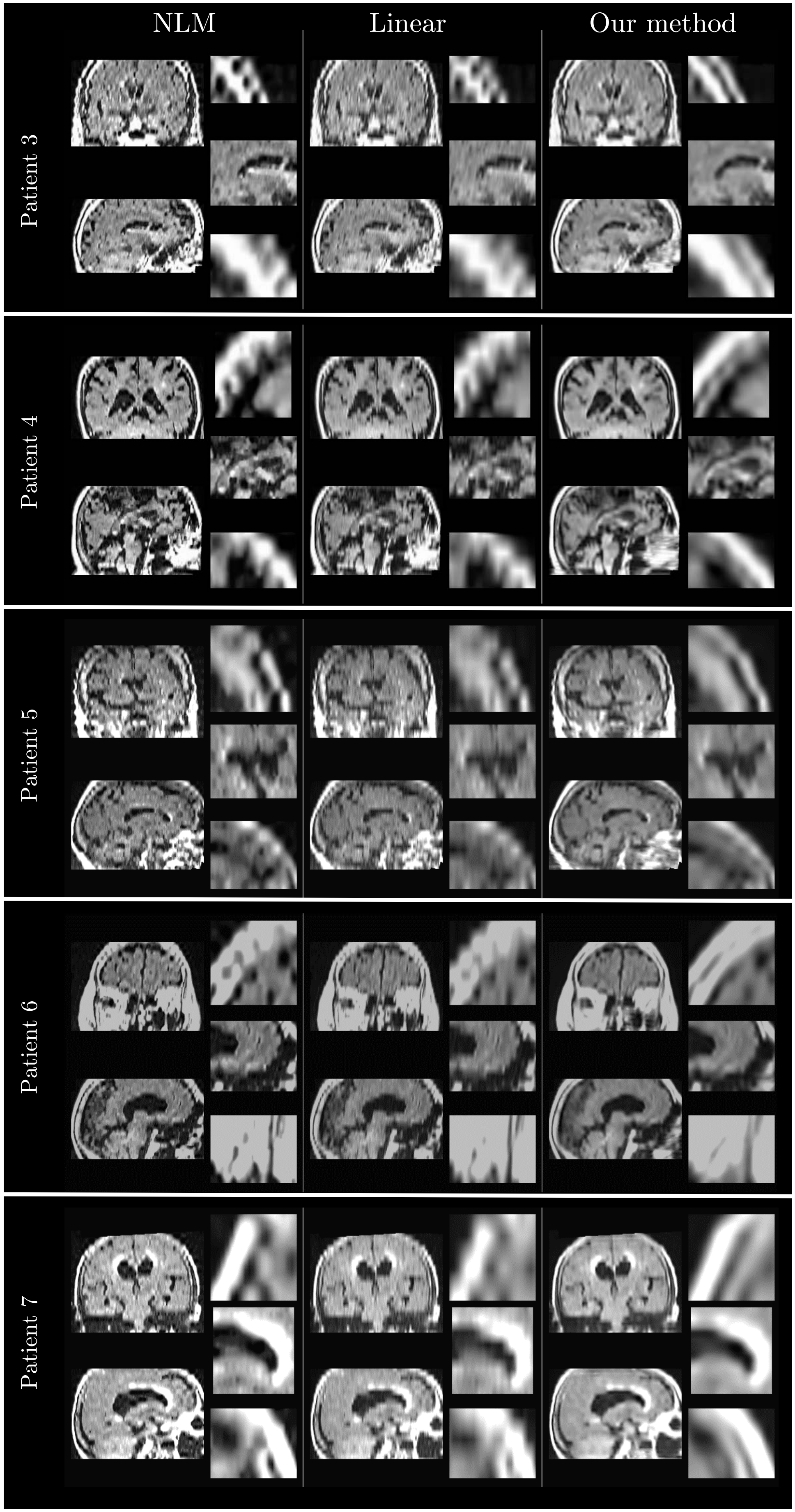

Fig. 12: Additional restorations in the clinical dataset. Reconstruction using NLM, linear interpolation and our method.